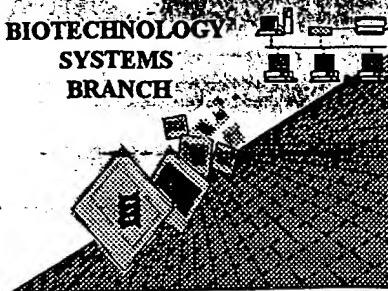


## **RAW SEQUENCE LISTING** **ERROR REPORT**

BIOTECHNOLOGY  
SYSTEMS  
BRANCH



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/323,597A

Source: 1642

Date Processed by STIC: 9/11/2000

RECEIVED

SEP 18 2000

TECH CENTER 1600/2900

**THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.**

**PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:**

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR FURTHER INFORMATION, PLEASE TELEPHONE MARK SPENCER, 703-308-4212.

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER VERSION 3.0 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

### **Checker Version 3.0**

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 - 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

**Checker Version 3.0 can be down loaded from the USPTO website at the following address:**

**<http://www.uspto.gov/web/offices/pac/checker>**

# Raw Sequence Listing Error Summary

## ERROR DETECTED SUGGESTED CORRECTION

SERIAL NUMBER:

09/323, 597A

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

RECEIVED

SEP 18 2000

TECH CENTER 1600/2900

- 1        Wrapped Nucleics      The number/text at the end of each line "wrapped" down to the next line.  
This may occur if your file was retrieved in a word processor after creating it.  
Please adjust your right margin to .3, as this will prevent "wrapping".
- 2        Wrapped Aminos      The amino acid number/text at the end of each line "wrapped" down to the next line.  
This may occur if your file was retrieved in a word processor after creating it.  
Please adjust your right margin to .3, as this will prevent "wrapping".
- 3        Incorrect Line Length      The rules require that a line not exceed 72 characters in length. This includes spaces.
- 4        Misaligned Amino Acid      The numbering under each 5th amino acid is misaligned. This may be caused by the use of tabs  
Numbering      between the numbering. It is recommended to delete any tabs and use spacing between the numbers.
- 5        Non-ASCII      This file was not saved in ASCII (DOS) text, as required by the Sequence Rules.  
Please ensure your subsequent submission is saved in ASCII text so that it can be processed.
- 6        Variable Length      Sequence(s)        contain n's or Xaa's which represented more than one residue.  
As per the rules, each n or Xaa can only represent a single residue.  
Please present the maximum number of each residue having variable length and  
indicate in the (ix) feature section that some may be missing.
- 7        PatentIn ver. 2.0 "bug"      A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid  
sequence(s)       . Normally, PatentIn would automatically generate this section from the  
previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section  
to the subsequent amino acid sequence. This applies primarily to the mandatory <220>-<223>  
sections for Artificial or Unknown sequences.
- 8        Skipped Sequences      Sequence(s)        missing. If intentional, please use the following format for each skipped sequence:  
(OLD RULES)      (2) INFORMATION FOR SEQ ID NO:X:  
                         (i) SEQUENCE CHARACTERISTICS:(Do not insert any headings under "SEQUENCE CHARACTERISTICS")  
                         (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X:  
                         This sequence is intentionally skipped  
  
Please also adjust the "(iii) NUMBER OF SEQUENCES:" response to include the skipped sequence(s).
- 9        Skipped Sequences      Sequence(s)        missing. If intentional, please use the following format for each skipped sequence.  
(NEW RULES)      <210> sequence id number  
                         <400> sequence id number  
                         000
- 10        Use of n's or Xaa's      Use of n's and/or Xaa's have been detected in the Sequence Listing.  
(NEW RULES)      Use of <220> to <223> is MANDATORY if n's or Xaa's are present.  
In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 11        Use of <213>Organism      Sequence(s)        are missing this mandatory field or its response.  
(NEW RULES)
- 12        Use of <220>Feature      Sequence(s)        are missing the <220>Feature and associated headings.  
(NEW RULES)      Use of <220> to <223> is MANDATORY if <213>ORGANISM is "Artificial" or "Unknown"  
Please explain source of genetic material in <220> to <223> section.  
(See "Federal Register," 6/01/98, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of new Rules)
- 13        PatentIn ver. 2.0 "bug"      Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted  
file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing).  
Instead, please use "File Manager" or any other means to copy file to floppy disk.

1642

RAW SEQUENCE LISTING      DATE: 09/11/2000  
 PATENT APPLICATION: US/09/323,597A      TIME: 09:07:15

Input Set : A:\seqlist.txt  
 Output Set: N:\CRF3\09112000\I323597A.raw

Does Not Comply  
 Corrected Diskette Needed

3 <110> APPLICANT: Daniel E. Afar  
 4      Rene S. Hubert  
 5      Kahan Leong  
 6      Arthur B. Raitano  
 7      Douglas C. Saffran  
 8      Stephen C. Mitchell  
 10 <120> TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS  
 11      AND THERAPY OF PROSTATE AND COLON CANCER  
 14 <130> FILE REFERENCE: 129.8USU1  
 16 <140> CURRENT APPLICATION NUMBER: 09/323,597A  
 17 <141> CURRENT FILING DATE: 1999-06-01  
 19 <150> PRIOR APPLICATION NUMBER: 60/087,598  
 20 <151> PRIOR FILING DATE: 1998-06-01  
 22 <150> PRIOR APPLICATION NUMBER: 60/091,474  
 23 <151> PRIOR FILING DATE: 1998-06-29  
 25 <150> PRIOR APPLICATION NUMBER: 60/129,521  
 26 <151> PRIOR FILING DATE: 1999-04-14  
 28 <160> NUMBER OF SEQ ID NOS: 13  
 30 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
 32 <210> SEQ ID NO: 1  
 33 <211> LENGTH: 1738  
 34 <212> TYPE: DNA  
 35 <213> ORGANISM: Homo sapiens  
 37 <220> FEATURE:  
 38 <221> NAME/KEY: CDS  
 39 <222> LOCATION: (112)...(1588)  
 41 <400> SEQUENCE: 1  
 42 ggcggaggcg gagcgaggagg ggcggggcg gggagcgccg cctggagcgc ggcaggtcat      60  
 43 attgaacatt ccagatacct atcattactc gatgctgttg ataacagcaa g atg gct      117  
 44      Met Ala  
 45      1  
 47 ttg aac tca ggg tca cca cca gct att gga cct tac tat gaa aac cat      165  
 48 Leu Asn Ser Gly Ser Pro Pro Ala Ile Gly Pro Tyr Tyr Glu Asn His  
 49      5      10      15  
 51 gga tac caa ccg gaa aac ccc tat ccc gca cag ccc act gtg gtc ccc      213  
 52 Gly Tyr Gln Pro Glu Asn Pro Tyr Pro Ala Gln Pro Thr Val Val Pro  
 53      20      25      30  
 55 act gtc tac gag gtg cat ccg gct cag tac tac ccg tcc ccc gtg ccc      261  
 56 Thr Val Tyr Glu Val His Pro Ala Gln Tyr Tyr Pro Ser Pro Val Pro  
 57      35      40      45      50  
 59 cag tac gcc ccg agg gtc ctg acg cag gct tcc aac ccc gtc gtc tgc      309  
 60 Gln Tyr Ala Pro Arg Val Leu Thr Gln Ala Ser Asn Pro Val Val Cys  
 61      55      60      65  
 63 acg cag ccc aaa tcc cca tcc ggg aca gtg tgc acc tca aag act aag      357  
 64 Thr Gln Pro Lys Ser Pro Ser Gly Thr Val Cys Thr Ser Lys Thr Lys  
 65      70      75      80  
 67 aaa gca ctg tgc atc acc ttg acc ctg ggg acc ttc ctc gtg gga gct      405

RAW SEQUENCE LISTING                      DATE: 09/11/2000  
 PATENT APPLICATION:    US/09/323,597A            TIME: 09:07:15

Input Set : A:\seqlist.txt  
 Output Set: N:\CRF3\09112000\I323597A.raw

```

68 Lys Ala Leu Cys Ile Thr Leu Thr Leu Gly Thr Phe Leu Val Gly Ala
69      85      90      95
71 gcg ctg gcc gct ggc cta ctc tgg aag ttc atg ggc agc aag tgc tcc      453
72 Ala Leu Ala Ala Gly Leu Leu Trp Lys Phe Met Gly Ser Lys Cys Ser
73      100      105      110
75 aac tct ggg ata gag tgc gac tcc tca ggt acc tgc atc aac ccc tct      501
76 Asn Ser Gly Ile Glu Cys Asp Ser Ser Gly Thr Cys Ile Asn Pro Ser
77 115      120      125      130
79 aac tgg tgt gat ggc gtg tca cac tgc ccc ggc ggg gag gac gag aat      549
80 Asn Trp Cys Asp Gly Val Ser His Cys Pro Gly Gly Glu Asp Glu Asn
81      135      140      145
83 cgg tgt gtt cgc ctc tac gga cca aac ttc atc ctt cag gtg tac tca      597
84 Arg Cys Val Arg Leu Tyr Gly Pro Asn Phe Ile Leu Gln Val Tyr Ser
85      150      155      160
87 tct cag agg aag tcc tgg cac cct gtg tgc caa gac gac tgg aac gag      645
88 Ser Gln Arg Lys Ser Trp His Pro Val Cys Gln Asp Asp Trp Asn Glu
89      165      170      175
91 aac tac ggg cgg gcg gcc tgc agg gac atg ggc tat aag aat aat ttt      693
92 Asn Tyr Gly Arg Ala Ala Cys Arg Asp Met Gly Tyr Lys Asn Asn Phe
93      180      185      190
95 tac tct agc caa gga ata gtg gat gac agc gga tcc acc agc ttt atg      741
96 Tyr Ser Ser Gln Gly Ile Val Asp Asp Ser Gly Ser Thr Ser Phe Met
97 195      200      205      210
99 aaa ctg aac aca agt gcc ggc aat gtc gat atc tat aaa aaa ctg tac      789
100 Lys Leu Asn Thr Ser Ala Gly Asn Val Asp Ile Tyr Lys Lys Leu Tyr
101      215      220      225
103 cac agt gat gcc tgt tct tca aaa gca gtg gtt tct tta cgc tgt ata      837
104 His Ser Asp Ala Cys Ser Ser Lys Ala Val Val Ser Leu Arg Cys Ile
105      230      235      240
107 gcc tgc ggg gtc aac ttg aac tca agc cgc cag agc agg att gtg ggc      885
108 Ala Cys Gly Val Asn Leu Asn Ser Ser Arg Gln Ser Arg Ile Val Gly
109      245      250      255
111 ggc gag agc gcg ctc ccg ggg gcc tgg ccc tgg cag gtc agc ctg cac      933
112 Gly Glu Ser Ala Leu Pro Gly Ala Trp Pro Trp Gln Val Ser Leu His
113      260      265      270
115 gtc cag aac gtc cac gtg tgc gga ggc tcc atc atc acc ccc gag tgg      981
116 Val Gln Asn Val His Val Cys Gly Gly Ser Ile Ile Thr Pro Glu Trp
117 275      280      285      290
119 atc gtg aca gcc gcc cac tgc gtg gaa aaa cct ctt aac aat cca tgg      1029
120 Ile Val Thr Ala Ala His Cys Val Glu Lys Pro Leu Asn Asn Pro Trp
121      295      300      305
123 cat tgg acg gca ttt gcg ggg att ttg aga caa tct ttc atg ttc tat      1077
124 His Trp Thr Ala Phe Ala Gly Ile Leu Arg Gln Ser Phe Met Phe Tyr
125      310      315      320
127 gga gcc gga tac caa gta gaa aaa gtg att tct cat cca aat tat gac      1125
128 Gly Ala Gly Tyr Gln Val Glu Lys Val Ile Ser His Pro Asn Tyr Asp
129      325      330      335
131 tcc aag acc aag aac aat gac att gcg ctg atg aag ctg cag aag cct      1173
132 Ser Lys Thr Lys Asn Asn Asp Ile Ala Leu Met Lys Leu Gln Lys Pro

```

DATE: 09/11/2000

TIME: 09:07:15

Output Set: N:\CRF3\09112000\I323597A.raw

OK

RAW SEQUENCE LISTING                      DATE: 09/11/2000  
 PATENT APPLICATION:    US/09/323,597A              TIME: 09:07:15

Input Set : A:\seqlist.txt  
 Output Set: N:\CRF3\09112000\I323597A.raw

194				100				105				110
195	Cys	Ser	Asn	Ser	Gly	Ile	Glu	Cys	Asp	Ser	Ser	Gly
196			115					120				125
197	Pro	Ser	Asn	Trp	Cys	Asp	Gly	Val	Ser	His	Cys	Pro
198			130				135				140	
199	Glu	Asn	Arg	Cys	Val	Arg	Leu	Tyr	Gly	Pro	Asn	Phe
200	145				150				155			
201	Tyr	Ser	Ser	Gln	Arg	Lys	Ser	Trp	His	Pro	Val	Cys
202					165				170			175
203	Asn	Glu	Asn	Tyr	Gly	Arg	Ala	Ala	Cys	Arg	Asp	Met
204				180					185			190
205	Asn	Phe	Tyr	Ser	Ser	Gln	Gly	Ile	Val	Asp	Asp	Ser
206			195				200				205	
207	Phe	Met	Lys	Leu	Asn	Thr	Ser	Ala	Gly	Asn	Val	Asp
208		210					215				220	
209	Leu	Tyr	His	Ser	Asp	Ala	Cys	Ser	Ser	Lys	Ala	Val
210	225				230					235		240
211	Cys	Ile	Ala	Cys	Gly	Val	Asn	Leu	Asn	Ser	Ser	Arg
212				245					250			255
213	Val	Gly	Gly	Glu	Ser	Ala	Leu	Pro	Gly	Ala	Trp	Pro
214			260					265				270
215	Leu	His	Val	Gln	Asn	Val	His	Val	Cys	Gly	Gly	Ser
216		275					280				285	
217	Glu	Trp	Ile	Val	Thr	Ala	Ala	His	Cys	Val	Glu	Lys
218		290					295				300	
219	Pro	Trp	His	Trp	Thr	Ala	Phe	Ala	Gly	Ile	Leu	Arg
220	305				310				315			320
221	Phe	Tyr	Gly	Ala	Gly	Tyr	Gln	Val	Glu	Lys	Val	Ile
222			325					330				335
223	Tyr	Asp	Ser	Lys	Thr	Lys	Asn	Asn	Asp	Ile	Ala	Leu
224			340					345				350
225	Lys	Pro	Leu	Thr	Phe	Asn	Asp	Leu	Val	Lys	Pro	Val
226		355					360				365	
227	Pro	Gly	Met	Met	Leu	Gln	Pro	Glu	Gln	Leu	Cys	Trp
228		370					375				380	
229	Gly	Ala	Thr	Glu	Glu	Lys	Gly	Lys	Thr	Ser	Glu	Val
230	385				390				395			400
231	Lys	Val	Leu	Leu	Ile	Glu	Thr	Gln	Arg	Cys	Asn	Ser
232			405						410			415
233	Asp	Asn	Leu	Ile	Thr	Pro	Ala	Met	Ile	Cys	Ala	Gly
234			420					425				430
235	Asn	Val	Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro
236		435					440				445	
237	Lys	Asn	Asn	Ile	Trp	Trp	Leu	Ile	Gly	Asp	Thr	Ser
238		450					455				460	
239	Cys	Ala	Lys	Ala	Tyr	Arg	Pro	Gly	Val	Tyr	Gly	Asn
240	465				470				475			480
241	Thr	Asp	Trp	Ile	Tyr	Arg	Gln	Met	Arg	Ala	Asp	Gly
242				485				490				

RAW SEQUENCE LISTING                      DATE: 09/11/2000  
 PATENT APPLICATION: US/09/323,597A        TIME: 09:07:15

Input Set : A:\seqlist.txt  
 Output Set: N:\CRF3\09112000\I323597A.raw

```

244 <210> SEQ ID NO: 3
245 <211> LENGTH: 2479
246 <212> TYPE: DNA
247 <213> ORGANISM: Homo sapiens
249 <220> FEATURE:
250 <221> NAME/KEY: CDS
251 <222> LOCATION: (57)...(1533)
253 <400> SEQUENCE: 3
254 gtcattattga acattccaga tacctatcat tactcgatgc tgttgataac agcaag atg   59
255                                     Met
256                                     1
258 gct ttg aac tca ggg tca cca cca gct att gga cct tac tat gaa aac   107
259 Ala Leu Asn Ser Gly Ser Pro Pro Ala Ile Gly Pro Tyr Tyr Glu Asn
260                                     5 10 15
262 cat gga tac caa ccg gaa aac ccc tat ccc gca cag ccc act gtg gtc   155
263 His Gly Tyr Gln Pro Glu Asn Pro Tyr Pro Ala Gln Pro Thr Val Val
264                                     20 25 30
266 ccc act gtc tac gag gtg cat ccg gct cag tac tac ccg tcc ccc gtg   203
267 Pro Thr Val Tyr Glu Val His Pro Ala Gln Tyr Tyr Pro Ser Pro Val
268                                     35 40 45
270 ccc cag tac gcc ccg agg gtc ctg acg cag gct tcc aac ccc gtc gtc   251
271 Pro Gln Tyr Ala Pro Arg Val Leu Thr Gln Ala Ser Asn Pro Val Val
272 50 55 60 65
274 tgc acg cag ccc aaa tcc cca tcc ggg aca gtg tgc acc tca aag act   299
275 Cys Thr Gln Pro Lys Ser Pro Ser Gly Thr Val Cys Thr Ser Lys Thr
276                                     70 75 80
278 aag aaa gca ctg tgc atc acc ttg acc ctg ggg acc ttc ctc gtg gga   347
279 Lys Lys Ala Leu Cys Ile Thr Leu Thr Leu Gly Thr Phe Leu Val Gly
280                                     85 90 95
282 gct gcg ctg gcc gct ggc cta ctc tgg aag ttc atg ggc agc aag tgc   395
283 Ala Ala Leu Ala Ala Gly Leu Leu Trp Lys Phe Met Gly Ser Lys Cys
284                                     100 105 110
286 tcc aac tct ggg ata gag tgc gac tcc tca ggt acc tgc atc aac ccc   443
287 Ser Asn Ser Gly Ile Glu Cys Asp Ser Ser Gly Thr Cys Ile Asn Pro
288                                     115 120 125
290 tct aac tgg tgt gat ggc gtg tca cac tgc ccc ggc ggg gag gac gag   491
291 Ser Asn Trp Cys Asp Gly Val Ser His Cys Pro Gly Gly Glu Asp Glu
292 130 135 140 145
294 aat cgg tgt gtt cgc ctc tac gga cca aac ttc atc ctt cag atg tac   539
295 Asn Arg Cys Val Arg Leu Tyr Gly Pro Asn Phe Ile Leu Gln Met Tyr
296                                     150 155 160
298 tca tct cag agg aag tcc tgg cac cct gtg tgc caa gac gac tgg aac   587
299 Ser Ser Gln Arg Lys Ser Trp His Pro Val Cys Gln Asp Asp Trp Asn
300                                     165 170 175
302 gag aac tac ggg cgg gcg gcc tgc agg gac atg ggc tat aag aat aat   635
303 Glu Asn Tyr Gly Arg Ala Ala Cys Arg Asp Met Gly Tyr Lys Asn Asn
304                                     180 185 190
306 ttt tac tct agc caa gga ata gtg gat gac agc gga tcc acc agc ttt   683
307 Phe Tyr Ser Ser Gln Gly Ile Val Asp Asp Ser Gly Ser Thr Ser Phe

```

<210> 5  
 <211> 388  
 <212> DNA  
 <213> Homo Sapiens

*See item 10 on Error Summary sheet*

<400> 5  
 gatcttcctg ctgagtcctt tccaggggcc aattttggat gagcatggag ctgtcacctc 60  
 tcagctgctg gatgacttga gatgaaaaag gagagacatg gaaagggaga cagccaggtg 120  
 gcacctgcag cggctgccct ctggggccac ttggtagtgt cccagccta cctctccaca 180  
 aggggatttt gctgatgggt tcttanagcc ttagcagccc tggatgggtg ccagaaataa 240  
 agggaccagc ccttcatggg tggtgacgtg gtaattcactt gtaaggggaa cagaaacatt 300  
 ttgtttctta tggggtgaga atatagacag tgcccttggg gcgaggggaag caattgaaaa 360  
 ggaacttgcc ctgagcactc ctggtgca 388



VERIFICATION SUMMARY

DATE: 09/11/2000

PATENT APPLICATION: US/09/323,597A

TIME: 09:07:16

Input Set : A:\seqlist.txt

Output Set: N:\CRF3\09112000\I323597A.raw

L:168 M:361 W: Invalid Split Codon, Sequence data for SEQ ID#: 1  
 L:379 M:361 W: Invalid Split Codon, Sequence data for SEQ ID#: 3  
 L:477 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:5  
 L:477 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:5  
 L:477 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:5  
 L:477 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:5  
 L:477 M:340 W: (46) "n" or "Xaa" used: Feature required, for SEQ ID#:5  
 L:478 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:5  
 L:478 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:5  
 L:478 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:5  
 L:478 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:5  
 M:340 Repeated in SeqNo=5